Measuring ISP Response to VeriSign Site Finder

Benjamin Edelman
Berkman Center for Internet & Society
Harvard Law School
Motivation

• Why measure generally
• Special implications of disabling Site Finder
  – If find Site Finder disabled in many places
  – If find Site Finder disabled in only a few places
Methodology I: Surveys / Form Submissions

• Problems
  – Burden on respondents
  – Representativeness
  – Accuracy

• Results - Submissions to date report blocking by
  – AOL
  – Earthlink
  – San Bernardino County School System
  – Time Warner Cable (partial?)
  – …
Methodology II: Site Finder Log Analysis

• Problem
  – Availability of log files for analysis
    • “[A]s a matter of privacy and operational security we do not provide our log files to any outside party”
      – Tom Galvin, VeriSign spokesman (email)

• Results
  – Unknown
Methodology III: Inference from selected user clickstreams

• Idea: Get data about selected users / networks accesses to Site Finder content, look for trends.

• Problems
  – Potential privacy concerns
  – Obtaining necessary data
  – Users with non-default nameservers
  – Users who request Site Finder content manually (other than by mistyping domain names)
  – Works best for large networks
Reference:
Inference from selected user clickstreams

see also

Technical Responses to Unilateral Internet Authority:
The Deployment of VeriSign “Site Finder” and ISP Response

http://cyber.law.harvard.edu/tlds/sitefinder
with Jonathan Zittrain
Intro to Alexa Toolbar

• Provides search shortcut, related links, etc.
• 10+ million downloads. Active users unknown.
• Representativeness
  – Users generally representative of Internet community.
  – Possible under-emphasis on technical community.
  – Possible over-emphasis on Southeast Asia.
Alexa Data Set

- Date through Sep. 29 as to requests for web pages on verisign.com by Alexa Toolbar users

- Data elements:
  - User IP address (/24)
  - Date & time
  - URL requested
Some Noteworthy Networks: China

Page Views (by Alexa Users)
Some Noteworthy Networks: Adelphia
Aggregation: Idealizations

- Idealization of traffic to date:
  - Introduction of Site Finder; jump in traffic.
  - Site Finder remains operational; traffic remains constant and high.

- Possible future trend in traffic:
  - Large ISPs disable Site Finder; traffic to Site Finder drops.
  - Introduction of Site Finder; jump in traffic.
Aggregation: Data (ordinary, public Alexa rank chart)
A page from a document containing a graph titled "Aggregation: Data (summed across all networks)". The graph shows page views by Alexa Users from various networks over a period from 9/15 to 9/29. There is a significant drop in traffic on 9/24 and 9/25.

Annotators note:
- Anomalous drop in traffic:
  - DOS attack?
  - Alexa data collection failure?
Additional Analysis with Add’l Data

• Possible data sources
  – Site Finder web server log files
  – Google Toolbar log files, other toolbars

• Could address Site Finder blocking
  – by smaller networks
  – by networks with fewer or no Alexa Toolbar users
Analyzing Page-Views on Site Finder

How many Site Finder users:

- Click on a “Did you mean?” link?
- Click on a “Popular Category”?
- Read the Terms of Use?
- Leave Site Finder without clicking on anything?
## Analyzing Page-Views on Site Finder

<table>
<thead>
<tr>
<th>URL</th>
<th>% of Accesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>sitefinder.verisign.com/lpc – default Site Finder page</td>
<td>81.58%</td>
</tr>
<tr>
<td>sitefinder.verisign.com/spc</td>
<td>16.19%</td>
</tr>
<tr>
<td>sitefinder.verisign.com</td>
<td>1.60%</td>
</tr>
<tr>
<td>sitefinder.verisign.com/help.jsp</td>
<td>0.18%</td>
</tr>
<tr>
<td>sitefinder.verisign.com/preferences.jsp</td>
<td>0.16%</td>
</tr>
<tr>
<td>sitefinder.verisign.com/terms.jsp</td>
<td>0.15%</td>
</tr>
<tr>
<td>sitefinder.verisign.com/privacy.jsp</td>
<td>0.11%</td>
</tr>
<tr>
<td>sitefinder.verisign.com/whatsthis.jsp</td>
<td>&lt;0.01%</td>
</tr>
<tr>
<td>sitefinder.verisign.com/pdf/sitefinderdevguide.pdf</td>
<td>&lt;0.01%</td>
</tr>
</tbody>
</table>

Note: No data available in Alexa logs as to clicks within “Did you mean?”
Session-Based Analysis

• “What proportion of users click on…” rather than “What proportion of page-views are for…”
• Difficult using Alexa Toolbar data because low order byte of IP address is unavailable.
  – 12% of user-sessions include at least one “Popular Category” view.
  – (“Session” = “request from same /24 within 5 minutes”)


Benjamin Edelman
http://cyber.law.harvard.edu/edelman
edelman@law.harvard.edu